Theoretical Computer Science 410 (2009) 997–1000

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Note

On the simplification of infinite morphic words

Juha Honkala

Department of Mathematics, University of Turku, 20014 Turku, Finland

ARTICLE INFO

Article history:
Received 25 July 2008
Received in revised form 6 November 2008
Accepted 14 December 2008
Communicated by D. Perrin

Keywords:
Infinite word
Morphic word
Elementary morphism

ABSTRACT

We study the simplification of infinite morphic words by using elementary morphisms. In particular, we give a new proof of a result of Cobham, stating that a morphic image of a morphic word is a finite word or a morphic word.

1. Introduction

Elementary morphisms introduced by Ehrenfeucht and Rozenberg [3] are a powerful tool in the study of free monoid morphisms. Ehrenfeucht and Rozenberg used elementary morphisms to give a very enlightening solution for the D0L sequence equivalence problem [4]. Linna has used them to solve the D0L prefix problem [11]. Elementary morphisms also give an easy solution to the problem of deciding whether or not an infinite pure morphic word is ultimately periodic [5,13]. For further applications of elementary morphisms see, for example, [7–10].

In this paper, we study the simplification of infinite morphic words by using elementary morphisms. An infinite word u is called pure morphic if u is generated by iterating a morphism. An infinite word v is called morphic if v is obtained by applying a coding to a pure morphic word. Cobham [2] has proved that a pure morphic word is always obtained by applying a coding to a pure morphic word generated by iterating a nonerasing morphism. Cobham has also observed that the morphic image of a morphic word is a morphic word (if not finite). Allouche and Shallit [1] give a detailed proof of this result. For another proof, see Pansiot [12]. Here we will use elementary morphisms to give new proofs for these important results.

We assume some familiarity with the basic properties of infinite morphic words, see [1].

2. Definitions and results

We use standard language-theoretic notation and terminology, see e.g. [1]. If u, v are words, we denote $u \le v$ if u is a prefix of v. If $w \in X^*$ is a nonempty word, then $\mathrm{alph}(w)$ is the set of all letters of X occurring in w. A mapping $g : X^* \longrightarrow Y^*$ is called a morphism if $g(uv) = g(u)g(v)$ for all $u, v \in X^*$.

Let $g : X^* \longrightarrow X^*$ be a morphism. If, for all $x \in X$, the letter x has an occurrence in $g(x)$, then g is called *cyclic*. Now let $g : X^* \longrightarrow Y^*$ be a morphism. Then g is *nonerasing* if $g(x)$ is a nonempty word for all $x \in X$. If $g(x) \in Y$ for all $x \in X$, then g is called a *coding*. Finally, g is called *elementary* if there do not exist a set Z and two morphisms $f_1 : X^* \longrightarrow Z^*, f_2 : Z^* \longrightarrow Y^*$ such that $g = f_2 f_1$ and $\mathrm{card}(Z) < \mathrm{card}(X)$.

Example 1. Let $X = \{a, b\}$. The Fibonacci morphism $f : X^* \longrightarrow X^*$ defined by $f(a) = ab, f(b) = a$ is elementary but not cyclic. The morphism $g : X^* \longrightarrow X^*$ defined by $g(a) = ab, g(b) = abab$ is cyclic but not elementary.

E-mail address: juha.honkala@utu.fi.

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2008.12.037

Let $g : X^* \longrightarrow X^*$ be a morphism and let $w \in X^*$ be a word. Suppose $g(w) = wu$ where $u \in X^*$ and suppose that $\{g^n(w) \mid n \geq 0\}$ is an infinite set. Then the infinite word $g^\omega(w)$ is defined by

$$g^\omega(w) = \lim_{n\to\infty} g^n(w) = wug(u)g^2(u)\ldots.$$

An infinite word w is called a *pure morphic word* if there exist a morphism $g : X^* \longrightarrow X^*$ and a letter $x \in X$ such that $g^\omega(x)$ exists and

$$w = g^\omega(x).$$

More generally, an infinite word w is called a *morphic word* if there exist a pure morphic word v over an alphabet X and a coding $h : X^* \longrightarrow Y^*$ such that

$$w = h(v).$$

In this paper, we will use elementary morphisms to prove the following result concerning the simplification of infinite morphic words.

Theorem 1. *Let $g : X^* \longrightarrow X^*$ and $h : X^* \longrightarrow Y^*$ be morphisms and let $w \in X^*$ be a word such that $g^\omega(w)$ exists. Assume that $hg^\omega(w)$ is an infinite word. Then there exist an alphabet X_1, a nonerasing morphism $g_1 : X_1^* \longrightarrow X_1^*$, a coding $h_1 : X_1^* \longrightarrow Y^*$ and a letter $x \in X_1$ such that $g_1^\omega(x)$ exists and*

$$hg^\omega(w) = h_1 g_1^\omega(x).$$

In particular, a morphic image of a morphic word is morphic (if not finite). Moreover, each pure morphic word can be obtained by applying a coding to a pure morphic word generated by iterating a nonerasing morphism.

Theorem 1 was first observed by Cobham [2]. For a proof, see also Pansiot [12] and Allouche, Shallit [1].

Below we will need the following lemma.

Lemma 1. *Let $g : X^* \longrightarrow X^*$ be an elementary morphism. Then there exists a positive integer n such that*

$$x \in \mathrm{alph}(g^n(x)) = \mathrm{alph}(g^{2n}(x))$$

for all $x \in X$.

Proof. By [8, Lemma 2] there is a positive integer t such that g^t is cyclic. (We can choose $t = \mathrm{card}(X)!$.) Hence

$$x \in \mathrm{alph}(g^{ti}(x)) \subseteq \mathrm{alph}(g^{t(i+1)}(x))$$

for all $x \in X$ and $i \geq 1$. This implies the claim. □

3. Proof of Theorem 1

In this section, we will prove Theorem 1 by using elementary morphisms. Similar ideas were used in [6] to study the simplification of HD0L power series. Theorem 1 and the main result of [6] are incompatible, in the sense that neither implies the other.

Consider an infinite word $hg^\omega(w)$. We first replace g by an elementary morphism.

Lemma 2. *Let $g : X^* \longrightarrow X^*$ and $h : X^* \longrightarrow Y^*$ be morphisms and let $w \in X^*$ be a word such that $g^\omega(w)$ exists. Then there exist an alphabet X_1, an elementary morphism $g_1 : X_1^* \longrightarrow X_1^*$, a morphism $h_1 : X_1^* \longrightarrow Y^*$ and a word $w_1 \in X_1^*$ such that $g_1^\omega(w_1)$ exists and*

$$hg^\omega(w) = h_1 g_1^\omega(w_1).$$

Proof. We use induction on the cardinality of X. If $\mathrm{card}(X) = 1$, then g is elementary and there is nothing to prove. Consider, now, an alphabet X and suppose that the claim holds for all smaller alphabets. If g is elementary, there again is nothing to prove. Assume that g is not elementary and let X_1 be an alphabet smaller than X such that there exist morphisms $f_1 : X^* \longrightarrow X_1^*$ and $f_2 : X_1^* \longrightarrow X^*$ with $g = f_2 f_1$. Define $f = f_1 f_2$ and $v = f_1(w)$. Then

$$f_2 f^n(v) = g^{n+1}(w)$$

for all $n \geq 0$. Hence the set $\{f^n(v) \mid n \geq 0\}$ is infinite. Because $w \leq f_2 f_1(w)$, we have $v = f_1(w) \leq f_1 f_2 f_1(w) = f_1 f_2(v) = f(v)$. These observations imply that $f^\omega(v)$ exists (because $f^n(v)$ is a prefix of $f^{n+1}(v)$ for all $n \geq 0$) and

$$hg^\omega(w) = hf_2 f^\omega(v).$$

Now the claim follows inductively. □

Consider again an infinite word $hg^\omega(w)$ and suppose that g is elementary. We next replace h and g by a nonerasing morphism and a cyclic morphism, respectively.

Lemma 3. *Let $g : X^* \longrightarrow X^*$ be an elementary morphism and let $w \in X^*$ be a word such that $g^\omega(w)$ exists. Let $h : X^* \longrightarrow Y^*$ be a morphism such that $hg^\omega(w)$ is an infinite word. Then there exist an alphabet X_1, a cyclic morphism $g_1 : X_1^* \longrightarrow X_1^*$, a nonerasing morphism $h_1 : X_1^* \longrightarrow Y^*$ and a word $w_1 \in X_1^*$ such that $g_1^\omega(w_1)$ exists and*

$$hg^\omega(w) = h_1 g_1^\omega(w_1).$$

Proof. By Lemma 1 we may assume that g is a morphism such that

$$x \in \mathrm{alph}(g(x)) = \mathrm{alph}(g^2(x)) \tag{1}$$

for all $x \in X$. (We no longer assume that g is elementary.)

Define

$$X_1 = \{x \in X \mid hg(x) \neq \varepsilon\}.$$

Define $\beta : X^* \longrightarrow X_1^*$ by

$$\beta(x) = \begin{cases} x & \text{if } x \in X_1 \\ \varepsilon & \text{otherwise} \end{cases}$$

and define $g_1 : X_1^* \longrightarrow X_1^*$ by

$$g_1(x) = \beta g(x)$$

for all $x \in X_1$. Then we have $g_1\beta(x) = \beta g(x)$ for all $x \in X$. This is clear if $x \in X_1$. Otherwise, $hg(x) = \varepsilon$ and $hg^2(x) = \varepsilon$, which implies that $\beta g(x) = \varepsilon$.

Next, define $h_1 : X_1^* \longrightarrow Y^*$ by

$$h_1(x) = hg(x)$$

for all $x \in X_1$ and define $w_1 = \beta(w)$. Then $hg = hg\beta$ and

$$hg^{n+1}(w) = hg\beta g^n(w) = hgg_1^n\beta(w) = h_1 g_1^n(w_1) \tag{2}$$

for all $n \geq 0$. Because $w \leq g(w)$, we have $w_1 = \beta(w) \leq \beta g(w) = g_1\beta(w) = g_1(w_1)$. This, together with (2), implies that $g_1^\omega(w_1)$ exists and

$$hg^\omega(w) = h_1 g_1^\omega(w_1).$$

Finally, g_1 is cyclic by (1) and h_1 is nonerasing by the definition of X_1. $\square$

Now consider an infinite word $hg^\omega(w)$ where h is nonerasing and g is cyclic. We next replace h and g by a coding and a cyclic morphism, respectively.

Lemma 4. *Let $g : X^* \longrightarrow X^*$ be a cyclic morphism and let $w \in X^*$ be a word such that $g^\omega(w)$ exists. Let $h : X^* \longrightarrow Y^*$ be a nonerasing morphism. Then there exist an alphabet X_1, a cyclic morphism $g_1 : X_1^* \longrightarrow X_1^*$, a coding $h_1 : X_1^* \longrightarrow Y^*$ and a word $w_1 \in X_1^*$ such that $g_1^\omega(w_1)$ exists and*

$$hg^\omega(w) = h_1 g_1^\omega(w_1).$$

Proof. Let

$$X_1 = \{(x, i) \mid x \in X, 1 \leq i \leq |h(x)|\}$$

be a new alphabet. Define the morphism $\alpha : X^* \longrightarrow X_1^*$ by

$$\alpha(x) = (x, 1) \ldots (x, |h(x)|)$$

for $x \in X$. Define $w_1 = \alpha(w)$. Let $g_1 : X_1^* \longrightarrow X_1^*$ be a cyclic morphism such that

$$g_1\alpha(x) = \alpha g(x)$$

for all $x \in X$. The existence of g_1 follows because g is cyclic. Let $h_1 : X_1^* \longrightarrow Y^*$ be the coding such that

$$h_1\alpha(x) = h(x)$$

for all $x \in X$. The existence of h_1 follows because for all $x \in X$ the length of $\alpha(x)$ equals the length of $h(x)$. Then we have

$$hg^n(w) = h_1\alpha g^n(w) = h_1 g_1^n\alpha(w) = h_1 g_1^n(w_1)$$

for all $n \geq 0$. Because $w \leq g(w)$, we have $w_1 = \alpha(w) \leq \alpha g(w) = g_1\alpha(w) = g_1(w_1)$. Hence $g_1^\omega(w_1)$ exists and

$$hg^\omega(w) = h_1 g_1^\omega(w_1). \quad \square$$

The construction used to prove Lemma 3 (resp. Lemma 4) is the same as the construction used to prove Lemma 7 (resp. Lemma 6) in [6]. (The proofs are more or less verbatim from [6].) However, in [6] all morphisms are monomial morphisms between semirings of polynomials in noncommuting variables. On the other hand, in [6] we do not have to consider the existence of infinite words defined by the morphisms.

Now the following simple lemma concludes the proof of Theorem 1.

Lemma 5. *Suppose $g : X^* \longrightarrow X^*$ is a nonerasing morphism and $w \in X^*$ is a word such that $g^\omega(w)$ exists. Then there is an alphabet X_1, a nonerasing morphism $g_1 : X_1^* \longrightarrow X_1^*$, a coding $h_1 : X_1^* \longrightarrow X^*$ and a letter $x \in X_1$ such that $g_1^\omega(x)$ exists and*

$$g^\omega(w) = h_1 g_1^\omega(x).$$

Proof. Let $w = a_1 \dots a_t$ where $a_j \in X$ for $j = 1, \dots, t$. Choose the smallest $s \le t$ such that $\{g^n(a_s) \mid n \ge 0\}$ is infinite. Then

$$g^\omega(w) = g^\omega(a_1 \dots a_s)$$

and $g(a_i) = a_i$ for $i = 1, \dots, s-1$. Let $g(a_s) = a_s a_{s+1} \dots a_q$ where again $a_j \in X$ for $j = s, \dots, q$. Let $g^\omega(w) = a_1 \dots a_s u$.

Next, assume that $s \ge 2$. Choose new letters $\ell_1, \dots, \ell_s$ and let $X_1 = X \cup \{\ell_1, \dots, \ell_s\}$. Let $g_1 : X_1^* \longrightarrow X_1^*$ be the extension of g such that $g_1(\ell_1) = \ell_1 \ell_2$, $g_1(\ell_i) = \ell_{i+1}$ $(i = 2, \dots, s-1)$ and $g_1(\ell_s) = a_{s+1} \dots a_q$. Then g_1 is nonerasing and

$$g_1^\omega(\ell_1) = \ell_1 \ell_2 \dots \ell_s u,$$

which implies the claim. □

References

[1] J.-P. Allouche, J. Shallit, Automatic Sequences, Theory, Applications, Generalizations, Cambridge University Press, Cambridge, 2003.
[2] A. Cobham, On the Hartmanis-Stearns problem for a class of tag machines, in: IEEE Conference Record of Ninth Annual Symposium on Switching and Automata Theory, 1968, pp. 51–60.
[3] A. Ehrenfeucht, G. Rozenberg, Simplifications of homomorphisms, Inform. Control 38 (1978) 298–309.
[4] A. Ehrenfeucht, G. Rozenberg, Elementary homomorphisms and a solution of the D0L sequence equivalence problem, Theoret. Comput. Sci. 7 (1978) 169–183.
[5] T. Harju, M. Linna, On the periodicity of morphisms on free monoids, Theoret. Inform. Appl. 20 (1986) 47–54.
[6] J. Honkala, On the simplification of HD0L power series, J. UCS 8 (2002) 1040–1046.
[7] J. Honkala, The equivalence problem for DF0L languages and power series, J. Comput. System Sci. 65 (2002) 377–392.
[8] J. Honkala, The equivalence problem of polynomially bounded D0L systems – a bound depending only on the size of the alphabet, Theory Comput. Systems 36 (2003) 89–103.
[9] J. Honkala, The D0L ω-equivalence problem, Internat. J. Found. Comput. Sci. 18 (2007) 181–194.
[10] J. Honkala, Cancellation and periodicity properties of iterated morphisms, Theoret. Comput. Sci. 391 (2008) 61–64.
[11] M. Linna, The decidability of the D0L prefix problem, Internat. J. Comput. Math. 6 (1977) 127–142.
[12] J.-J. Pansiot, Hiérarchie et fermeture de certaines classes de tag-systèmes, Acta Inform. 20 (1983) 179–196.
[13] J.-J. Pansiot, Decidability of periodicity for infinite words, Theoret. Inform. Appl. 20 (1986) 43–46.